Optimal control theory and static optimization in economics

DANIEL LEONARD

NGO VAN LONG

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http: //www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http: //www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1992
Reprinted 1993, 1994, 1995, 1996, 1998

Typeset in Times

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data is available

ISBN 0-521-33158-7 hardback
ISBN 0-521-33746-1 paperback

Transferred to digital printing 2002

Contents

CHAPTER 1

Static optimization

In this chapter we deal with problems involving the choice of values for a finite number of variables in order to maximize some objective. Sometimes the values the variables may take are unrestricted; at other times they are restricted by equality constraints and also by inequality constraints. In the course of the presentation an important class of functions will emerge; they are called *concave functions* and are closely associated with "nice" maximum problems. They will be encountered throughout this book. For this reason we weave the concept of concavity of functions through the exposition of maximization problems. This is done to suit our purposes, but concave functions have other important properties in their own right.

The notation we use is fairly standard. If in doubt, the reader should refer to the appendix to this chapter, which also contains a reminder of the basic notions of multivariate calculus and some matrix algebra needed to follow the exposition.

1.1 Unconstrained optimization, concave and convex functions

In what follows we assume all functions to have continuous second-order derivatives, unless otherwise stated. Strictly speaking, all domains of definitions should be open subsets of the multidimensional real space so that no boundary problems arise.

1.1.1 *Unconstrained maximization*

Consider the problem of finding a set of values $x_1, x_2, \ldots, x_n$ to maximize the function $f(x_1, \ldots, x_n)$. We often write this as

$$\underset{\mathbf{x}}{\text{Maximize}}\ f(\mathbf{x}), \tag{1.1}$$

where $\mathbf{x}$ is understood to be an n-dimensional vector. We refer to the problem of (1.1) as an unconstrained maximum because no restrictions are placed on $\mathbf{x}$.

Necessary conditions. Suppose we find a solution to this problem and denote the optimal vector by $\mathbf{x}^*$. Consider an arbitrarily small deviation from $\mathbf{x}^*$, say $d\mathbf{x}$. If we have a maximum at $\mathbf{x}^*$, then f must not increase for any $d\mathbf{x}$.

The change in f is approximated by

$$df = \sum_i f_{x_i}(\mathbf{x}^*)\, dx_i.$$

Clearly, $df \leq 0$ if we have a maximum at $\mathbf{x}^*$. Furthermore, suppose we found some $d\mathbf{x}$ vector such that $df < 0$; then by using the deviation $(-d\mathbf{x})$ we would obtain an increase in f. Therefore, it must be that for any $d\mathbf{x}$ vector, df is equal to zero. The only way this can be achieved for arbitrary deviations is to require each derivative $f_{x_i}(\mathbf{x}^*)$ to vanish. Formally,

$$f(\mathbf{x}) \text{ reaches a maximum at } \mathbf{x}^* \text{ implies } f_{x_i}(\mathbf{x}^*) = 0, \quad i = 1, \ldots, n. \tag{1.2}$$

This is called the *first-order condition*. Several remarks must now be made. First, the above reasoning, hence (1.2), also applies to minimization problems. Second, we have been lax in defining a maximum. We should have distinguished a global maximum from a local maximum. We say that $f(x)$ reaches a *global maximum* at $\mathbf{x}^*$ if $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x}$ on its domain of definition (assumed to be an open set). We say $f(\mathbf{x})$ reaches a *local maximum* at $\mathbf{x}^*$ if $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x}$ "close" to $\mathbf{x}^*$ (i.e., for all $\mathbf{x}$ within δ units of distance from $\mathbf{x}^*$, where δ is some positive number). The local maximum is a much weaker concept than the global one. However, because our argument relies on arbitrarily small deviations from $\mathbf{x}^*$, it applies to both cases. The first-order condition (1.2) follows from the existence of a maximum; hence, it is a necessary condition for a maximum, but it is not the only one, as we now show. As we noted previously, condition (1.2) is necessary for a local minimum as well. The following condition, called the *second-order necessary condition,* takes a different form for a maximum than for a minimum.

To establish it we must take a Taylor's expansion (with remainder) of the function f about the point $\mathbf{x}^*$:

$$\begin{aligned} f(\mathbf{x}^* + d\mathbf{x}) = f(\mathbf{x}^*) &+ \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^*)(dx_i) \\ &+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}^*)(dx_i)(dx_j) + \cdots + R, \end{aligned} \tag{1.3a}$$

or in vector notation (see the Appendix for details),

$$f(\mathbf{x}^* + d\mathbf{x}) = f(\mathbf{x}^*) + (d\mathbf{x})' \cdot f_{\mathbf{x}}(\mathbf{x}^*) + \tfrac{1}{2}(d\mathbf{x})' \cdot f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \cdot (d\mathbf{x}) + \cdots + R, \tag{1.3b}$$

where $d\mathbf{x}$ is small enough (i.e., $\|d\mathbf{x}\| < \delta$) that higher-order terms vanish relative to second-order terms.

Suppose again that we have a (at least local) maximum, that is, $f(\mathbf{x}^*) \geq f(\mathbf{x}^*+d\mathbf{x})$, $\forall d\mathbf{x}$, $\|d\mathbf{x}\| < \delta$. Then $f_{\mathbf{x}}(\mathbf{x}^*) = 0$, and neglecting terms higher than the second order we have

$$\begin{aligned} f(\mathbf{x}^*+d\mathbf{x}) - f(\mathbf{x}^*) &= \tfrac{1}{2}(d\mathbf{x})' \cdot f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \cdot d\mathbf{x} \\ &\leq 0, \quad \text{because } \mathbf{x}^* \text{ is a maximum.} \end{aligned}$$

Since $(d\mathbf{x})' \cdot f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \cdot (d\mathbf{x})$ is negative or zero for all small deviation vectors $d\mathbf{x}$, the Hessian matrix of f evaluated at $\mathbf{x}^*$ must be negative-semidefinite. This is the *second-order necessary condition:*

$$f(\mathbf{x}) \text{ reaches a maximum at } \mathbf{x}^* \text{ implies } f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \text{ is negative-semidefinite.} \tag{1.4}$$

Again, (1.4) applies to global as well as local maxima.

Sufficient conditions (for a local maximum). It is unfortunately not possible to state conditions that are both necessary and sufficient for a function to reach a maximum. We can, however, easily provide sufficient conditions:

$$\begin{aligned} &\text{If } f_{x_i}(\mathbf{x}^*) = 0,\ i = 1, \ldots, n, \text{ and } f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \text{ is negative-definite,} \\ &\quad \text{then } f(\mathbf{x}) \text{ reaches a local maximum at } \mathbf{x}^*. \end{aligned} \tag{1.5}$$

To prove this we shall consider again Taylor's expansion in (1.3) and let $d\mathbf{x} \to 0$, so that the second-degree term dominates those of higher order while the first-degree term vanishes; we obtain $f(\mathbf{x}^*+d\mathbf{x}) < f(\mathbf{x}^*)$, thus establishing $\mathbf{x}^*$ as a local maximum.

1.1.2 *Global results and concave functions*

When we seek a maximum in an economic problem, it is most often a global one. Indeed, it is little comfort to know that we are doing the best we can but only if considering policies which differ minutely from the current one (local optimum). It is also clear that we will not be able to characterize a global maximum with conditions on the values of the function and its derivatives at the maximum itself; we will need to place restrictions on the overall shape of the function, restrictions that apply everywhere on the domain of definition, which we denote by X.

Consider the exact form of Taylor's expansion to the second degree: there exists a point $\mathbf{x}_t$ on the line segment between $\mathbf{x}$ and $\bar{\mathbf{x}}$ such that

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + (\mathbf{x}-\bar{\mathbf{x}})' \cdot f_{\mathbf{x}}(\bar{\mathbf{x}}) + \tfrac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})' \cdot \mathbf{H}(\mathbf{x}_t) \cdot (\mathbf{x}-\bar{\mathbf{x}}), \tag{1.6}$$

where $\mathbf{H}(\mathbf{x}_t)$ denotes the Hessian matrix of f, evaluated at the point $\mathbf{x}_t$. If we were to restrict our attention to functions with a negative-semidefinite matrix everywhere on its domain of definition, then the last term of (1.6) would be guaranteed to be nonpositive for any $\mathbf{x}_t$ and the requirement that $\bar{\mathbf{x}}$ be a global maximum (i.e., $f(\mathbf{x})-f(\bar{\mathbf{x}})\le 0\ \forall \mathbf{x}\in X$) would be equivalent to the first-order condition $f_{\mathbf{x}}(\bar{\mathbf{x}})=0$. We now formalize this argument.

Definition 1.1.1. A function with continuous second-order derivatives defined on a convex set X is concave if and only if its Hessian matrix is negative-semidefinite everywhere on its domain of definition X.

Theorem 1.1.1. Let $f(\mathbf{x})$ be a concave function; then it reaches a global maximum at $\bar{\mathbf{x}}$ if and only if $f_{\mathbf{x}}(\bar{\mathbf{x}})=0$.

Definition 1.1.1 applies only to functions with continuous second-order derivatives. It is useful to have a more general definition of concavity that does not require this assumption.

Definition 1.1.2. A function $f(\mathbf{x})$ with continuous first-order derivatives defined on a convex set X is concave if and only if

$$f(\mathbf{x}_2)-f(\mathbf{x}_1)\le(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_1),$$

for all $\mathbf{x}_1,\mathbf{x}_2$ on X.

Note that Definition 1.1.2 is less stringent than Definition 1.1.1 in terms of differentiability restrictions, since it requires continuity only for the first derivatives; this is the only difference between the two definitions. Indeed, if we assume that the function has continuous second-order derivatives, we can see that the two definitions are equivalent simply by writing down the exact form of Taylor's expansion. Given two arbitrary points $\mathbf{x}_1$ and $\mathbf{x}_2$, there exists a point $\mathbf{x}_t$ between them such that

$$f(\mathbf{x}_2)=f(\mathbf{x}_1)+(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_1)+\tfrac{1}{2}(\mathbf{x}_2-\mathbf{x}_1)'\cdot\mathbf{H}(\mathbf{x}_t)\cdot(\mathbf{x}_2-\mathbf{x}_1),$$

$$f(\mathbf{x}_2)-f(\mathbf{x}_1)-(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_1)=\tfrac{1}{2}(\mathbf{x}_2-\mathbf{x}_1)'\cdot\mathbf{H}(\mathbf{x}_t)\cdot(\mathbf{x}_2-\mathbf{x}_1)\le 0.$$

The geometric interpretation is simply that a tangent plane to the graph of $f(\mathbf{x})$ must remain everywhere above the graph, the equation for the tangent plane at $\mathbf{x}_1$ being

$$y=f(\mathbf{x}_1)+(\mathbf{x}-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_1).$$

This is illustrated in Figure 1.1a for functions of one variable. Definition 1.1.2 does not cover functions that have "kinks" and as such are not differentiable everywhere. To admit this case, a more general definition is needed.

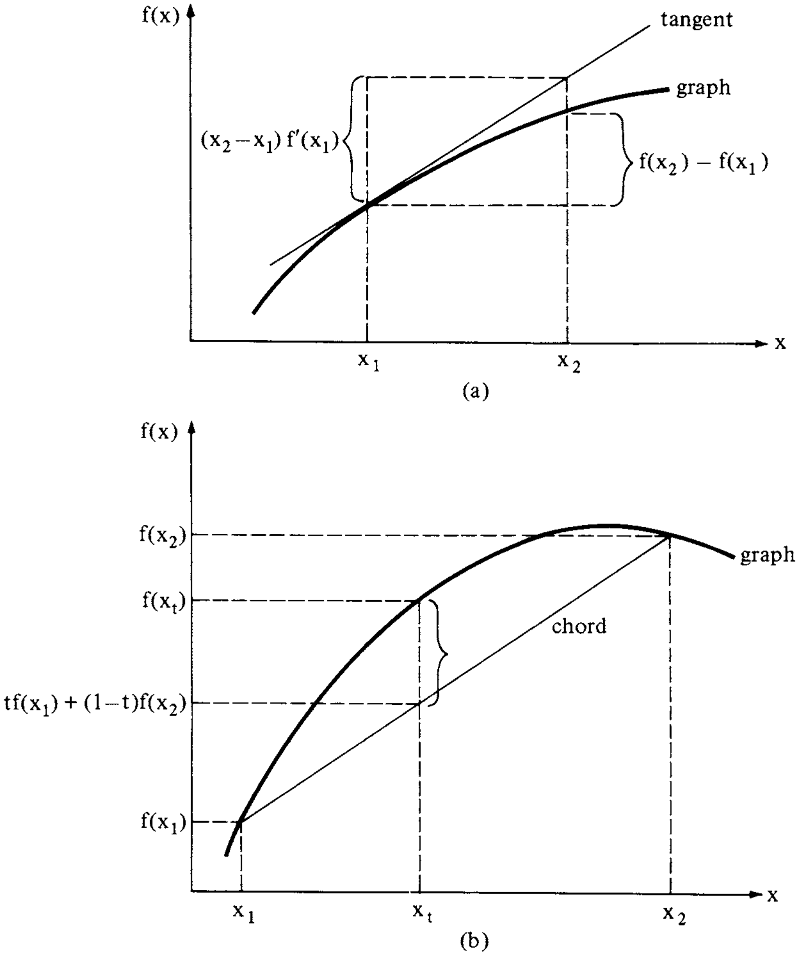

Figure 1.1

Definition 1.1.3. A function $f(\mathbf{x})$ defined on a convex set X is concave if and only if

$$f(\mathbf{x}_t) \geq tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2), \ 0 \leq t \leq 1, \text{ all } \mathbf{x}_1, \mathbf{x}_2 \text{ in } X,$$
$$\text{where } \mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_2.$$

If a function satisfies Definition 1.1.2, it also satisfies Definition 1.1.3. To see this we state Definition 1.1.2 in two instances:

$$f(\mathbf{x}_2)-f(\mathbf{x}_t)\le(\mathbf{x}_2-\mathbf{x}_t)'\cdot f_{\mathbf{x}}(\mathbf{x}_t)$$

and

$$f(\mathbf{x}_1)-f(\mathbf{x}_t)\le(\mathbf{x}_1-\mathbf{x}_t)'\cdot f_{\mathbf{x}}(\mathbf{x}_t),$$

where

$$\mathbf{x}_t=t\mathbf{x}_1+(1-t)\mathbf{x}_2 \text{ for some } t, \quad 0\le t\le 1.$$

Since

$$\mathbf{x}_2-\mathbf{x}_t=t(\mathbf{x}_2-\mathbf{x}_1) \quad\text{and}\quad \mathbf{x}_1-\mathbf{x}_t=-(1-t)(\mathbf{x}_2-\mathbf{x}_1),$$

we have

$$f(\mathbf{x}_2)-f(\mathbf{x}_t)\le t(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_t),$$

$$f(\mathbf{x}_1)-f(\mathbf{x}_t)\le -(1-t)(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_t).$$

Multiplying the first inequality by $(1-t)$, the second by t, and adding yields (with $0\le t\le 1$)

$$tf(\mathbf{x}_1)+(1-t)f(\mathbf{x}_2)-f(\mathbf{x}_t)\le 0,$$

which was to be proved.

Note that no differentiability properties are required in Definition 1.1.3. The geometric interpretation of this definition is that a line (or chord) joining two points of the graph always lies below the graph, since the left-hand side of the inequality represents the value of f at a convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ and the right-hand side is the same convex combination of the values of the function at $\mathbf{x}_1$ and $\mathbf{x}_2$ – hence the height of the point on the chord above $\mathbf{x}_t$. This is illustrated in Figure 1.1b for functions of one variable.

Concave functions have many notable properties; Theorem 1.1.2 lists some of the most useful ones.

Theorem 1.1.2

(i) Let $f(\mathbf{x})$ be a concave function and $k\ge 0$ a constant; then $kf(\mathbf{x})$ is a concave function.

(ii) Let $f(\mathbf{x})$ and $g(\mathbf{x})$ be concave functions; then $f(\mathbf{x})+g(\mathbf{x})$ is itself a concave function.

(iii) Let $f(\mathbf{x})$ be a concave function; then the upper contour set defined by $B(\bar{\mathbf{x}})\equiv\{\mathbf{x}\in R_n \mid f(\mathbf{x})\ge f(\bar{\mathbf{x}})\}$ is a convex set.

(The converse of (iii) is *not true*!)

The proofs of these results are straightforward; for instance, (iii) requires that we show that if $f(\mathbf{x}_1)\ge f(\bar{\mathbf{x}})$ and $f(\mathbf{x}_2)\ge f(\bar{\mathbf{x}})$, it follows that $f(\mathbf{x}_t)\ge f(\bar{\mathbf{x}})$; this is obvious from Definition 1.1.3.

Strictly concave functions: unique global maximum. While concave functions have the property that a solution of the first-order condition yields a global maximum, this does not ensure the uniqueness of that solution: a concave function may reach its global maximum at several points. For example, the following function is concave, but the first-order condition admits as a solution any point between 1 and 2; thus, the function reaches a global maximum at any x^* such that $1 \leq x^* \leq 2$.

$$f(x) = \begin{cases} x - 0.5x^2, & x < 1, \\ 0.5, & 1 \leq x \leq 2, \\ (x-1) - 0.5(x-1)^2, & 2 < x. \end{cases}$$

Other examples will be encountered in Section 1.1.5.

It is sometimes desirable to place more restrictions on the function so that if a maximum exists, it is the unique global maximum. We use this as a means of introducing a subclass of concave functions called *strictly concave functions*. Definitions 1.1.2 and 1.1.3 are adapted by simply requiring strict inequalities.

Definition 1.1.3′. A function $f(\mathbf{x})$ defined on a convex set X is strictly concave if and only if

$$f(\mathbf{x}_t) > tf(\mathbf{x}_1) + (1-t)f(\mathbf{x}_2), \quad 0 < t < 1,$$

for all $\mathbf{x}_1, \mathbf{x}_2$ in X, where $\mathbf{x}_1 \neq \mathbf{x}_2$ and $\mathbf{x}_t = t\mathbf{x}_1 + (1-t)\mathbf{x}_2$.

Definition 1.1.2′. A function $f(\mathbf{x})$ with continuous first-order derivatives defined on a convex set X is strictly concave if and only if

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) < (\mathbf{x}_2 - \mathbf{x}_1)' \cdot f_{\mathbf{x}}(\mathbf{x}_1)$$

for all $\mathbf{x}_1$ and $\mathbf{x}_2$ in X, where $\mathbf{x}_1 \neq \mathbf{x}_2$.

It is obvious from Definition 1.1.2′ that $f_{\mathbf{x}}(\mathbf{x}_1) = 0$ is necessary and sufficient for $\mathbf{x}_1$ to be the unique global maximum of that function f.

We cannot claim that functions with continuous second-order derivatives are strictly concave if and only if their Hessian matrix is negative-definite, because some strictly concave functions have a Hessian matrix which becomes negative-semidefinite at some points. One instance is $f(x_1, x_2) = -(x_1)^4 - (x_2)^2$, which is negative-definite everywhere but at $x_1 = 0$, when it is negative-semidefinite. We must be content with the following theorem.

Theorem 1.1.3. A function that is defined on a convex set X and has a negative-definite Hessian matrix everywhere on X is strictly concave.

The reader is invited to prove this result using Definition 1.1.2′.

1.1.3 *Unconstrained minimization and convex functions*

Results for minimization problems are just mirror images of those for maximization problems and are obtained by replacing $f(\mathbf{x})$ by $-f(\mathbf{x})$. Thus, the *first-order necessary condition for a local minimum at* $\mathbf{x}^*$ is

$$f_i(\mathbf{x}^*)=0, \quad i=1,\dots,n, \tag{1.7}$$

and the *second-order necessary condition* is

$$f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \text{ is positive-semidefinite.} \tag{1.8}$$

The *sufficient conditions for a local minimum at* $\mathbf{x}^*$ are

$$f_{\mathbf{x}}(\mathbf{x}^*)=0 \quad \text{and} \quad f_{\mathbf{x}\mathbf{x}'}(\mathbf{x}^*) \text{ is positive-definite.} \tag{1.9}$$

Similarly, we have to define convex functions in order to obtain global results on minimization. Corresponding to Definitions 1.1.1, 1.1.2, and 1.1.3 we now have the following (results on strictly convex functions are indicated in parentheses).

Definition 1.1.4. A function with continuous second-order derivatives defined on a convex set is (strictly) convex if and only if its Hessian matrix is positive-semidefinite (if its Hessian matrix is positive-definite).

Definition 1.1.5. A function $f(\mathbf{x})$ with continuous first-order derivatives defined on a convex set X is (strictly) convex if and only if

$$f(\mathbf{x}_2)-f(\mathbf{x}_1)\geq(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_1) \quad \text{for all } \mathbf{x}_1,\mathbf{x}_2 \text{ in } X,$$

$$(f(\mathbf{x}_2)-f(\mathbf{x}_1)>(\mathbf{x}_2-\mathbf{x}_1)'\cdot f_{\mathbf{x}}(\mathbf{x}_1) \quad \text{for all } \mathbf{x}_1,\mathbf{x}_2 \text{ in } X, \text{ where } \mathbf{x}_1\neq\mathbf{x}_2).$$

Definition 1.1.6. A function $f(\mathbf{x})$ defined on a convex set X is (strictly) convex if and only if

$$f(\mathbf{x}_t)\leq tf(\mathbf{x}_1)+(1-t)f(\mathbf{x}_2), \quad 0\leq t\leq 1, \text{ all } \mathbf{x}_1,\mathbf{x}_2 \text{ in } X,$$

$$(f(\mathbf{x}_t)<tf(\mathbf{x}_1)+(1-t)f(\mathbf{x}_2), \quad 0<t<1, \text{ all } \mathbf{x}_1,\mathbf{x}_2 \text{ in } X, \ \mathbf{x}_1\neq\mathbf{x}_2).$$

Theorem 1.1.4

(i) Let $f(\mathbf{x})$ be a convex function and $k\geq 0$ a constant; then $kf(\mathbf{x})$ is a convex function.
(ii) Let $f(\mathbf{x})$ and $g(\mathbf{x})$ be convex functions; then $f(\mathbf{x})+g(\mathbf{x})$ is itself a convex function.
(iii) Let $f(\mathbf{x})$ be a convex function; then the lower contour set defined by $W(\bar{\mathbf{x}})\equiv\{\mathbf{x}\in R_n \mid f(\mathbf{x})\leq f(\bar{\mathbf{x}})\}$ is a convex set. (The converse of (iii) is *not* true!)
(iv) Let $f(\mathbf{x})$ be a (strictly) convex function; then $-f(\mathbf{x})$ is a (strictly) concave function.

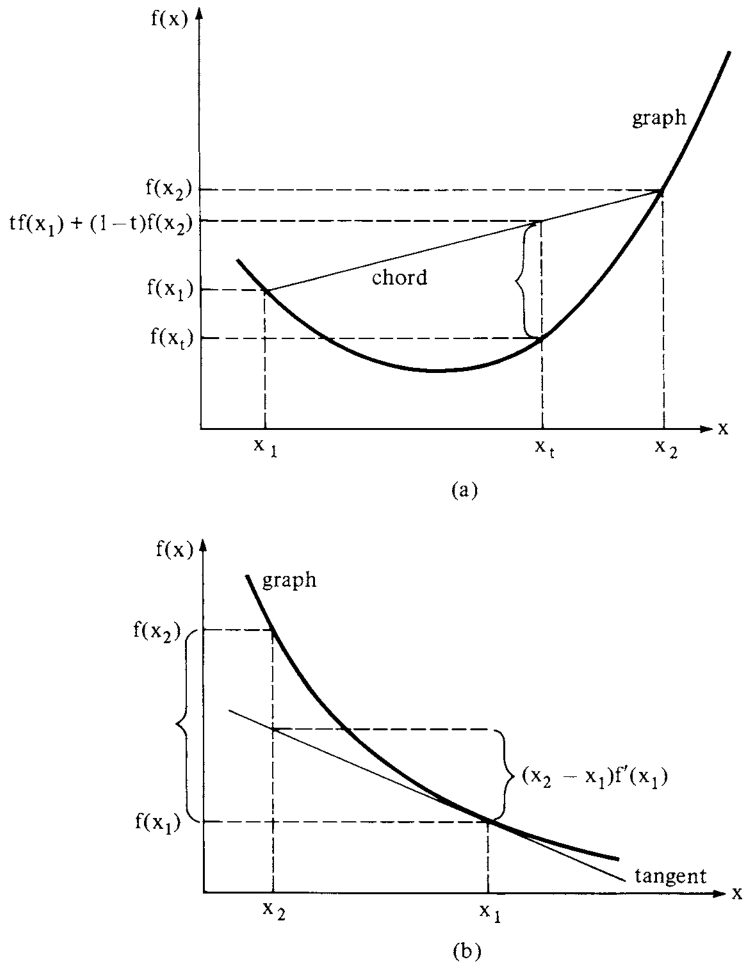

Figure 1.2

(v) A linear function is both convex and concave but not strictly either.

Definitions 1.1.5 and 1.1.6 are illustrated in Figure 1.2 for convex functions of one variable.

1.1.4 *Geometric representation*

Figures 1.3a and 1.3b represent the graphs of a concave and a convex function, respectively. It is important to realize that a concave function

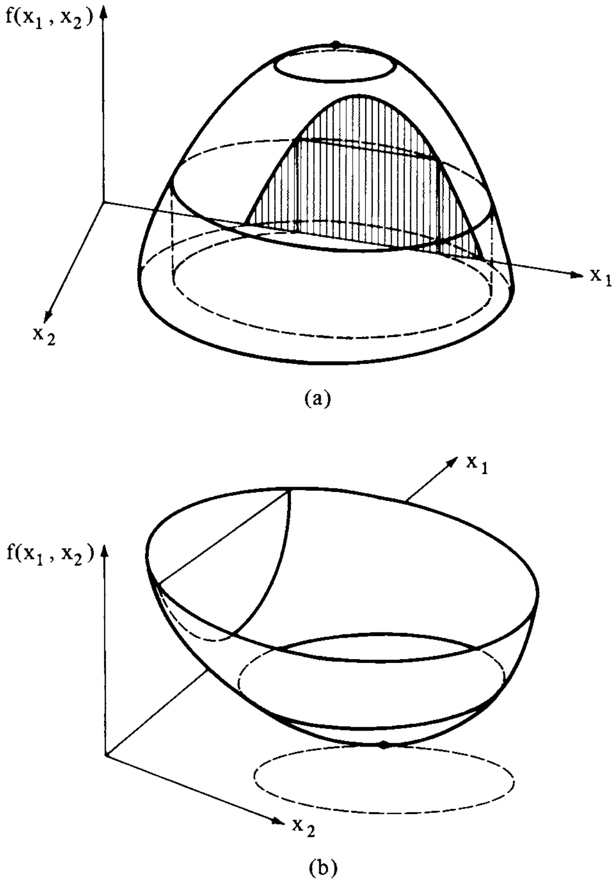

Figure 1.3

need not have a maximum, nor a convex function a minimum. If they do, then one is somewhat dome-shaped and the other bowl-shaped. It is then obvious that a rod connecting two points of the dome remains under it (Definition 1.1.3), while such a rod connecting two points of the bowl remains above its walls (Definition 1.1.6). It is clearly inconvenient to rely on three-dimensional diagrams; instead, we most often use level curves. We know that if a function is concave, its upper contour sets are convex sets. We use this information in Figure 1.4a to draw some level curves of a concave function, where the arrows indicate directions of increase of the function and one convex upper contour set is hatched. We can also verify that Definition 1.1.3 is satisfied: the function takes on the value c at points A and B; thus, it takes on a higher value at a point between

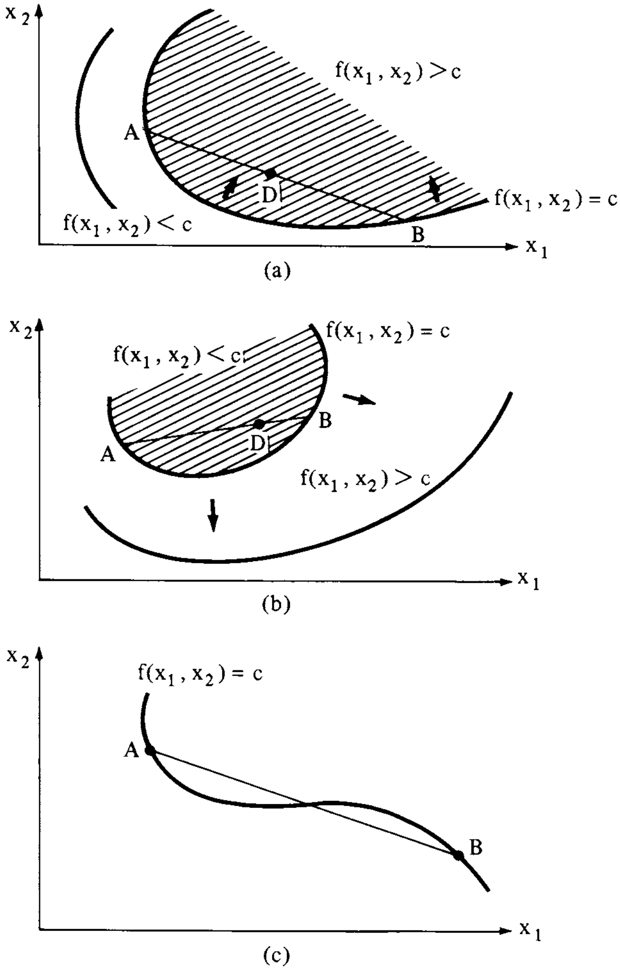

Figure 1.4

them, D, which is naturally within the convex upper contour set. A similar picture emerges for a convex function in Figure 1.4b, where a *lower* contour set is hatched and the arrows indicate directions of increase of the function. Finally note that a contour curve such as the one in Figure 1.4c cannot correspond to a concave (or a convex) function since it delineates no convex set on either side of it.

A word of warning is in order. Because concave functions have convex upper contour sets but some other functions do too, we cannot rely on this contour curve representation to *characterize concave functions* exactly. For many purposes, however, it will be adequate.

1.1.5 *Numerical examples and some useful functional forms*

It is useful to develop some "feel" for the concavity properties of functions so as to avoid always running back to the definitions. The knowledge of a few simple functions along with the composition rules already outlined and some more to follow is very helpful. We first list a few functions and the conditions for their concavity and/or convexity. The reader is invited to check these as exercises, using mainly Definitions 1.1.1 and 1.1.4.

$$f(\mathbf{x}) = \prod_{i=1}^{n} (x_i)^{\alpha_i} \text{ is concave for } x \geq 0 \text{ if and only if } \alpha_i \geq 0,\ \forall i, \text{ and } \sum_{i=1}^{n} \alpha_i \leq 1. \tag{1.10}$$

$$f(\mathbf{x}) = (a_0 + a_1 x_1 + \cdots + a_n x_n)^{\alpha}, \text{ defined when } a_0 + a_1 x_1 + \cdots + a_n x_n > 0 \text{ is concave if and only if } 0 \leq \alpha \leq 1; \text{ it is convex if and only if } \alpha \geq 1 \text{ or } \alpha \leq 0. \tag{1.11}$$

$$f(\mathbf{x}) = \mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x} \text{ is concave if and only if } \mathbf{A} \text{ is negative-semidefinite; it is convex if and only if } \mathbf{A} \text{ is positive-semidefinite.} \tag{1.12}$$

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \ln(x_i + a_i) \text{ is concave whenever it is defined (i.e., } x_i + a_i > 0, \text{ all } i) \text{ if and only if } \alpha_i \geq 0,\ \forall i; \text{ it is similarly convex if and only if } \alpha_i \leq 0,\ \forall i. \tag{1.13}$$

Theorem 1.1.5. An increasing concave function of concave functions is concave.

Proof. Let

$$W(\mathbf{x}^1, \ldots, \mathbf{x}^n) \equiv V(U^1(\mathbf{x}^1), \ldots, U^n(\mathbf{x}^n)),$$

where $\mathbf{x}^i$ denotes a vector of arbitrary dimension, V is increasing and concave in all U^i jointly, and U^i is concave in $\mathbf{x}^i$, $\forall i$. We use the standard notation for convex combinations: $\mathbf{z}_t \equiv t\mathbf{z}_1 + (1-t)\mathbf{z}_2$, $0 \leq t \leq 1$,

$$W(\mathbf{x}_t^1, \ldots, \mathbf{x}_t^n) = V(U^1(\mathbf{x}_t^1), \ldots, U^n(\mathbf{x}_t^n))$$
$$\geq V(tU^1(\mathbf{x}_1^1) + (1-t)\,U^1(\mathbf{x}_2^1), \ldots, tU^n(\mathbf{x}_1^n) + (1-t)\,U^n(\mathbf{x}_2^n)),$$

because all U^i are concave and V is increasing,

$$\geq tV(U^1(\mathbf{x}_1^1), \ldots, U^n(\mathbf{x}_1^n)) + (1-t)\,V(U^1(\mathbf{x}_2^1), \ldots, U^n(\mathbf{x}_2^n)),$$

by the concavity of V,

$$= tW(\mathbf{x}_1^1, \dots, \mathbf{x}_1^n) + (1-t)W(\mathbf{x}_2^1, \dots, \mathbf{x}_2^n). \quad \square$$

Theorem 1.1.6. Let $f(\mathbf{x})$ be a function of n variables and let $\mathbf{z} = -\mathbf{x}$ and $h(\mathbf{z}) \equiv f(\mathbf{x})$; then if $f(\mathbf{x})$ is concave (convex), so is $h(\mathbf{z})$.

The proof is obvious using, for instance, Definition 1.1.3. As an example, $f(x) = 1 - e^{-x}$ is concave in x; hence, $h(z) = 1 - e^{z}$ is concave in z, where $z = -x$.

We now consider a few numerical examples that may or may not possess a global maximum.

Example 1.1.1. Let $f(\mathbf{x}) = \mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x} + \mathbf{a}' \cdot \mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} -1 & 0.5 \\ 0.5 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{a} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}.$$

It is concave since $\mathbf{A}$ is negative-definite and the linear term does not affect concavity. To find a maximum, set the first-order derivatives to zero and solve: $f_1 = -2x_1 + x_2 - 1 = 0$ and $f_2 = x_1 - 2x_2 + 5 = 0$ yield $x_1 = 1$, $x_2 = 3$, the point at which f reaches its global maximum.

As we mentioned earlier, a function may reach its global maximum at many points; that is, the solution may not be unique. This is illustrated in the following example of a concave but not strictly concave function.

Example 1.1.2. Let $f(\mathbf{x}) = (x_1)^{0.3}(x_2)^{0.7} - 0.3x_1 - 0.7x_2$. We know that this function is defined and concave for all $\mathbf{x}$ positive (e.g., use (1.10)):

$$f_1 = 0.3(x_1)^{-0.7}(x_2)^{0.7} - 0.3 = 0,$$
$$f_2 = 0.7(x_1)^{0.3}(x_2)^{-0.3} - 0.7 = 0.$$

These first-order conditions have many solutions; namely, any $\mathbf{x}$ satisfying $x_1 = x_2$ is a solution. The global maximum value of f is zero and the upper part of its graph is shaped like the inside of a tunnel.

Example 1.1.3: saddle point. In this example we emphasize the idea that a function may be concave in all its variables but not necessarily concave in those variables jointly. We also introduce the concept of a saddle point. The example involves the function $f(x_1, x_2) = -(x_1)^2 + ax_1x_2 - (x_2)^2$ for various values of a.

Case (a). Let $f(x_1, x_2) = -(x_1)^2 - (x_2)^2$; then

$$\mathbf{H} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix},$$

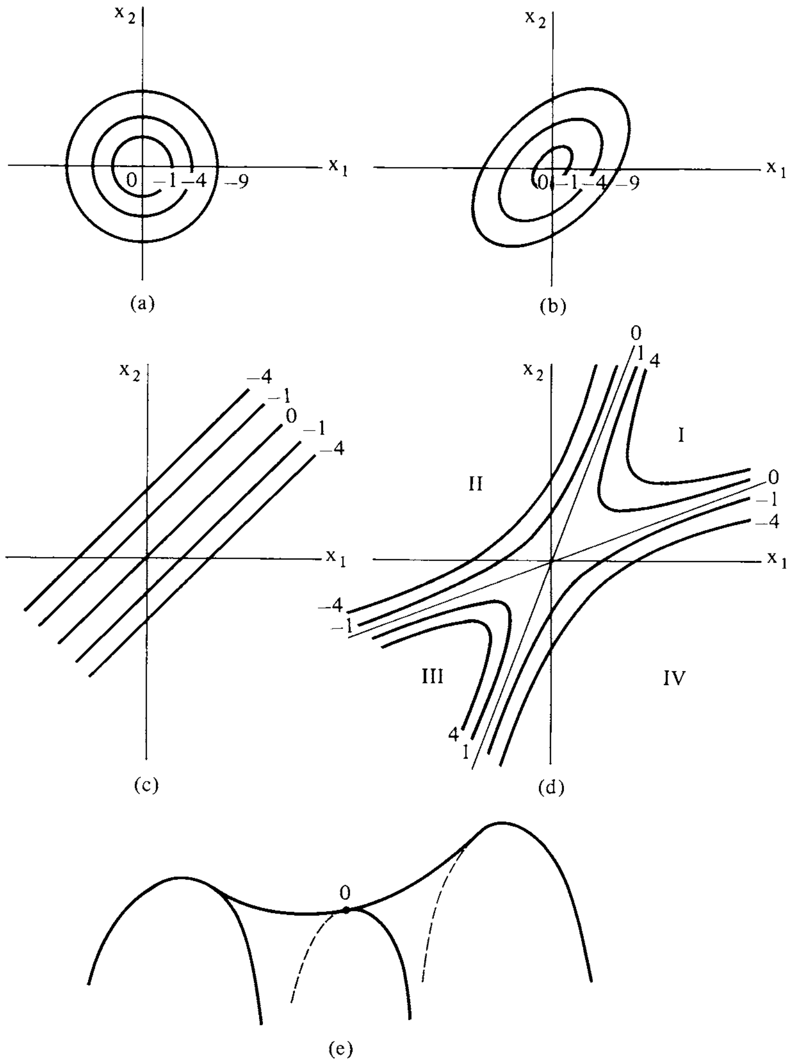

Figure 1.5

the function is concave, and $(0,0)$ is the global maximum. This is illustrated in Figure 1.5a. We proceed to "stretch" this function by introducing ever-increasing mixed terms.

Case (b). Let $f(x_1,x_2) = -(x_1)^2 + x_1x_2 - (x_2)^2$; then

$$\mathbf{H} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix},$$

the function is still concave, and $(0,0)$ is still a global maximum. The stretching is shown in Figure 1.5b.

Case (c). Let $f(x_1,x_2)=-(x_1)^2+2x_1x_2-(x_2)^2$; then

$$\mathbf{H}=\begin{bmatrix}-2 & 2\\ 2 & -2\end{bmatrix},$$

the function is still concave, but the first-order conditions only imply $x_1=x_2$; thus, there are many points at which the function reaches a global maximum. We again have a tunnel shape: the stretching has been carried out to an extent that we have a tubular shape with a horizontal top line. Note also that $|\mathbf{H}|=0$. Further stretching will destroy concavity, as we now see.

Case (d). Let $f(x_1,x_2)=-(x_1)^2+3x_1x_2-(x_2)^2$; then

$$\mathbf{H}=\begin{bmatrix}-2 & 3\\ 3 & -2\end{bmatrix};$$

the function is no longer concave in (x_1,x_2) because $|\mathbf{H}|=-5$, although it is still concave in x_1 and x_2 individually. The solution of the first-order conditions still is $(0,0)$, but we can no longer claim that it is a maximum. It is not a minimum either, but what we call a *saddle point.* The level curves are drawn in Figure 1.5d; the two straight lines corresponding to $f=0$ delineate four regions, and when we move from region I to III the origin appears to be a minimum, but when we cross the origin while moving from region II to IV it appears as a maximum. This is the essential property of a saddle point configuration: it appears as a maximum in some directions and as a minimum in others. These directions need not be the axes as Figure 1.5d shows. Thus if we cross the origin following any one axis, it appears as a maximum with respect to that variable, which is as it should be since f is concave in x_1 and concave in x_2, separately. We have drawn a three-dimensional representation of the graph in Figure 1.5e. The additional mixed term has lifted the ends of the tunnel; it does look something like a saddle. A mountain pass is another, less common description.

1.1.6 *Some economic applications*

We are now able to tackle any economic problem in which the objective is to maximize some objective and where the entities to be chosen are many while their choice is unrestricted. One such problem is *profit maximization* by a competitive firm, to which we now turn.

Let $f(x_1,\dots,x_n)$ be the output obtainable from input levels $x_1,\dots,x_n$. If output price is p, the price of input i is w_i, and some fixed cost is k, the maximization of profit reduces to choosing $(x_1,\dots,x_n)$ to maximize

$$pf(x_1, \ldots, x_n) - \sum_{i=1}^{n} w_i x_i - k.$$

If we assume that f is concave, the global maximum will be the solution of the n equations

$$pf_i - w_i = 0, \quad i = 1, \ldots, n.$$

The first term is the rate of increase in output per unit of input i at the margin (called the marginal physical product of input i) multiplied by the output price; this is called the marginal value product of input i (MVP_i for short). The above condition equates it to the price of input i; thus, the price of the input is equal to the contribution to revenue made by the marginal unit. This seems sensible, yet fails to relate maximization of profit to the concavity of the profit expression. We now seek to clarify this relationship. In general, economic sense dictates that if $\mathrm{MVP}_i > w_i$, we would gain by increasing the input level; conversely, $\mathrm{MVP}_i < w_i$ would lead us to decrease input. Suppose now that f is concave and indeed that $f_{ii} < 0$ for all i; then the derivative of f with respect to x_i decreases when x_i increases; hence, if x_i were to rise above the level $\mathbf{x}^*$ indicated by the first-order condition, w_i would exceed MVP_i and we would bring x_i back down. Similar reasoning shows that if x_i strays below that level, we should bring it back up. If, in contrast, f_{ii} were positive at $\mathbf{x}^*$, that point could not be a maximum, for an increase in x_i from $\mathbf{x}^*$ would increase MVP_i above w_i and induce further increases in x_i. Indeed, with a strictly convex production function we could reach an arbitrarily large profit; in other words, the problem would have no solution. This possibility should always be kept in mind for any problem in analytical economics, since we work with unspecified functional forms and a precise solution is never derived. As a way of illustrating this point we consider profit maximization when the production function is homogeneous.

Homogeneous production functions and returns to scale. Suppose $q = f(\mathbf{x})$ is a production function that is homogeneous of degree h (see the Appendix for definitions). From a starting point of $\mathbf{x}$ units of input, suppose that we scale the operations up by a factor of $t > 1$, that is, employ $t\mathbf{x}$ units of input; we will obtain an output $f(t\mathbf{x}) = (t)^h f(\mathbf{x})$, and hence we will have scaled up output by a factor $(t)^h$. Depending on the value of h, this factor $(t)^h$ will be larger or smaller than t and output will increase more or less than the input vector. More precisely,

$h < 1 \to (t)^h < t$: $f(\mathbf{x})$ exhibits decreasing returns to scale.

$h = 1 \to (t)^h = t$: $f(\mathbf{x})$ exhibits constant returns to scale.

$h > 1 \to (t)^h > t$: $f(\mathbf{x})$ exhibits increasing returns to scale.

Theorem 1.1.7. Let $f(\mathbf{x})$ be homogeneous of degree h, positively valued, and concave; then $0 \le h \le 1$.

Proof. Since f_j is homogeneous of degree $(h-1)$, Euler's theorem yields

$$\sum_i x_i f_{ij} = (h-1) f_j.$$

Multiplying by x_j, summing, and applying Euler's theorem again yields

$$\sum_j \sum_i x_i x_j f_{ij} = (h-1) \sum_j x_j f_j = (h-1)hf.$$

The left-hand side is the quadratic form $\mathbf{x}'\mathbf{H}\mathbf{x}$, where $\mathbf{H}$ is the Hessian matrix of f. Concavity of f ensures that it is nonpositive; hence, $f > 0$ implies $h(h-1) \le 0$. □

Note that our argument does not establish that if f is homogeneous of degree h and positively valued, then it is concave if and only if $h \le 1$. This is because on the left-hand side of the preceding equation the values x_i and x_j are from the same vector at which f_{ij} is evaluated, a weaker requirement than Definition 1.1.1 of concavity. As a counterexample consider the function $f(x_1, x_2) = (x_1^2 + x_2^2)^{1/4}$, $x_1 > 0$, $x_2 > 0$. It is positively valued and homogeneous of degree $\frac{1}{2}$ but not concave since its upper contour sets are clearly not convex sets.

We are now ready to examine the implications of alternative assumptions regarding the degree of homogeneity of the production function on the profit of the firm. In what follows we assume that all $\mathbf{x}$ are positively valued and $f(\mathbf{x}) > 0$, unless otherwise indicated.

If the profit expression $\pi = pf(\mathbf{x}) - \mathbf{w}' \cdot \mathbf{x}$ has an unconstrained maximum, it will satisfy the necessary conditions

$$pf_i(\mathbf{x}) = w_i, \quad i = 1, \ldots, n. \tag{1.14}$$

If $\mathbf{x}^*$ solves equation (1.14), multiplying by x_i^*, summing, and applying Euler's theorem yields

$$p \sum_i x_i^* f_i(\mathbf{x}^*) = \sum_i w_i x_i^*,$$

$$hpf(\mathbf{x}^*) = \mathbf{w}' \cdot \mathbf{x}^*.$$

Substituting in the profit expression, we get

$$\pi = (1-h) pf(\mathbf{x}^*).$$

Therefore, at $\mathbf{x}^*$, profit will be positive, zero, or negative, depending on whether there exist decreasing, constant, or increasing returns to scale, respectively. In the case of increasing returns, first note that the objective function cannot be concave (if it were, h could not exceed 1, by Theorem

1.1.7) and Theorem 1.1.1 fails us. Furthermore, as in the proof of Theorem 1.1.7, we can show that

$$\mathbf{x}^{*\prime}\cdot\mathbf{H}(\mathbf{x}^*)\cdot\mathbf{x}^* = (h-1)hf(\mathbf{x}^*) > 0, \quad \text{since } h>1.$$

This demonstrates that $\mathbf{H}(\mathbf{x}^*)$ is not negative-semidefinite and violates the second-order necessary condition for a maximum. Let us remark that increasing returns to scale are often associated with unbounded profit and as such are not consistent with the hypothesis of a price-taking firm. The case of decreasing returns poses no special problems, since we can assume that f is concave, but the case of constant returns to scale is more difficult to handle, although the profit expression is concave under the additional assumption of concavity for f. The problem is with the first-order conditions (1.14):

$$pf_i(\mathbf{x}) = w_i, \quad i=1,\ldots,n.$$

Recall that under constant returns, f_i is homogeneous of degree 0; therefore, if a vector $\mathbf{x}$ satisfies these conditions, so will any vector $t\cdot\mathbf{x}$, $t \geq 0$. The profit made with any of these vectors remains zero. The scale of operations is thus indeterminate and profit nil. This defect becomes a virtue when in some general equilibrium models such as those of international trade the focus is on the performance of each industry and the number and size of firms in each industry are not a matter of concern. There is, however, a further difficulty with the constant returns to scale assumption for an individual competitive firm. The problem is that for an arbitrary set of prices, equation (1.14) usually does not admit a solution, as we now demonstrate. Let $f(\mathbf{x})$ and $\mathbf{w}$ be fixed throughout, and suppose that at some price p^* (1.14) admits a vector $\mathbf{x}^*$ as a solution; then $\pi(\mathbf{x}^*) = 0$ and $t\mathbf{x}^*$ is also a solution, $t \geq 0$; this is a global maximum, since we assumed π to be concave. Now consider another output price, say p; the profit expression can be written as

$$\begin{aligned}\pi &= pf(\mathbf{x}) - \mathbf{w}'\cdot\mathbf{x}\\ &= (p-p^*)f(\mathbf{x}) + [p^*f(\mathbf{x}) - \mathbf{w}'\cdot\mathbf{x}].\end{aligned} \tag{1.15}$$

We know that the second term in (1.15) has a global maximum of zero at $\mathbf{x}^*$ (and at $t\mathbf{x}^*$), but if $p > p^*$ we can make the first term infinitely large by increasing t; hence, there is no maximum, and the first-order necessary conditions (1.14) do not hold anywhere (if they did, a global maximum would exist by concavity of π). Note that an arbitrary $\mathbf{x}$ value may well make profit negative even in this case. Conversely, suppose that $p < p^*$; then the second term has a global maximum of zero at $t\mathbf{x}^*$, but the first term can be only negative or zero. Hence, the maximum is found at the lower bound $\mathbf{x} = \mathbf{0}$ (i.e., $t = 0$), but this is not an unconstrained maximum and again the necessary first-order conditions (1.14) fail to have a solution. Note that when $p < p^*$, any vector $\mathbf{x} > \mathbf{0}$ yields a negative profit.

In order to get a more intuitive grasp of these results, consider a firm with two inputs. The equations (1.14) are $pf_1(x_1, x_2) = w_1$ and $pf_2(x_1, x_2) = w_2$. However, if f is homogeneous of degree 1, then f_1 and f_2 are homogeneous of degree 0. Consequently, these derivatives are simply functions of a single argument x_2/x_1 (the only one that matters, since scale is irrelevant), and both determine a value for it; unless the prices are in a particular configuration, these values will differ and no solution exists. The exact relationship is that output price p be equal to the unit cost function $c(1, w)$; see the definition of cost functions in Section 1.2.3. Let us now briefly illustrate these results with a numerical example.

Example 1.1.4

$$f(x_1, x_2) = 2(x_1)^{1/2}(x_2)^{1/2}, \quad w_1 = 1, \; w_2 = 2.$$

Equations (1.14) are

$$p(x_1)^{-1/2}(x_2)^{1/2} = 1 \quad \text{and} \quad p(x_1)^{1/2}(x_2)^{-1/2} = 2,$$

or

$$x_1/x_2 = p^2 \quad \text{and} \quad x_1/x_2 = 4/p^2.$$

Therefore, (1.14) is satisfied if and only if $p = p^* = \sqrt{2}$; then the optimal input mix is $x_1 = 2x_2$, the scale is arbitrary, and profit is zero, a global maximum. If, however, $p < p^*$, say $p = 1$, then $\pi = 2(x_1)^{1/2}(x_2)^{1/2} - x_1 - 2x_2$ and letting $(x_1/x_2)^{1/2} = u$, $\pi = x_2[-u^2 + 2u - 2]$. The bracketed expression is always negative and so is profit. Finally, if $p > p^*$, say $p = 2$, then $\pi = x_2[-u^2 + 4u - 2]$. This bracketed expression reaches a positive maximum of 2 when $u = 2$, that is, $x_1 = 4x_2$, and by letting x_2 be large we can generate arbitrarily large profits. Finally, note that an arbitrary choice of u may generate a negative profit, for example, $u = 4$, even with $p^* < p = 2$.

To gain some geometric insight into the matter, try to visualize the graph of a linearly homogeneous function of two variables. Because of the property $f(tx_1, tx_2) = tf(x_1, x_2)$ we see that the graph is "ruled from the origin"; a half-line from the origin to any point of the graph lies on the graph in its entirety. Visualize now the graph of input costs $C = w_1x_1 + w_2x_2$; it is a plane going through the origin. Let us now draw the graph of $pf(x_1, x_2)$ for low p values; it lies entirely below the cost plane: profit is everywhere negative. As p rises, the graph comes into contact with the plane, but it does so along an entire half-line from the origin. At